

Simultaneously calibrating solids, sugars and acidity of tomato products using PLS2 and NIR spectroscopy

André M.K. Pedro^{a,b}, Márcia M.C. Ferreira^{b,*}

^a Unilever Brasil Ltda, Av. Invernada 401, Valinhos, SP 13271-450, Brazil

^b Universidade Estadual de Campinas (UNICAMP), Chemistry Institute, Physical Chemistry Department, Campinas, SP 13084-971, Caixa Postal 6154, Brazil

Received 20 October 2006; received in revised form 14 March 2007; accepted 16 March 2007

Available online 24 March 2007

Abstract

In this work, the development of a robust spectroscopic procedure for determining, simultaneously and non-destructively, relevant quality parameters of processed tomato products (total and soluble solids, total acidity, total sugars, glucose and fructose), is described.

Samples of tomato concentrate products with total solids content ranging from 6.9 to 35.9% were collected from Latin America, the US and Europe and NIR spectra were acquired in the 4000–10,000 cm^{-1} region. The original spectra were pre-processed by mean-smoothing or by Fourier filter, followed by multiplicative signal correction (MSC) or derivatives. Partial least squares (PLS2 and PLS1) models were built and their predictive abilities were compared through the RMSEP of external validation.

The PLS2 regression had better predictive abilities for four out of the six properties under study, namely total solids, total sugars, glucose and fructose. Besides, the model was less complex than the PLS1 models in the sense that only four factors were demanded whilst from 4 to 11 factors were necessary for building the PLS1 models. The standard error of prediction (SEP%) of the PLS2 model for each property was: total solids, 2.67; soluble solids, 1.14; total acidity, 9.60; total sugar, 18.69; glucose, 11.60; and fructose, 13.45.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Fourier filter; Multivariate calibration; Tomato quality; Pre-processing; Convolution function

1. Introduction

Tomato is considerable relevant in the global agricultural market. It is the second most consumed vegetable worldwide, the first being potato [1]. In 2005, more than 30 million tonnes of tomato *in natura* were produced [2], from which about 85% were used for manufacturing industrialised products as tomato paste and sauces [3].

The fruit is mainly composed of water, tomato solids—soluble and insoluble, organic acids (mostly citric acid) and micronutrients such as carotenoids and vitamins A and C. Soluble solids are mainly sugars (sucrose and fructose) and salts, and are traditionally expressed as degrees Brix ($^{\circ}\text{Brix}$) [1,4,5]. Insoluble solids are mainly constituted of fibres, like cellulose and pectin. Usually, excluding seeds and skin, tomato presents

4.5–8.5% of total solids, depending on the variety, soil and climate conditions [1].

While tomato solids are relevant because they dictate the factory yield—the higher the amount of tomato solids, the less fruit is needed to produce processed tomato products, the relative amounts of sugars and acids define the taste of the final concentrate tomato product, thus largely determining consumer preference [1,4].

Despite the relevance of these attributes for the food industry, their quantification is still time- and labour-consuming: soluble solids are easily determined by refractometry, but total solids are quantified by oven drying under vacuum, a procedure which takes about 4 h. Total acidity is usually determined by titration using phenolphthalein as indicator, but the accurate determination of the end-point is difficult due to the tomato colour. Sugars are individually determined by HPLC, a technique that demands a considerable amount of organic solvents. In addition to that, the analysis is time- and labour-consuming in the sense that a thorough cleaning step is needed prior to the analysis [1,4–6].

* Corresponding author. Tel.: +55 19 3521 3102; fax: +55 19 3521 3023.

E-mail addresses: andre.k.pedro@unilever.com (A.M.K. Pedro), marcia@iqm.unicamp.br (M.M.C. Ferreira).

For solving analytical problems like the ones stated above, chemists have allied fast and non-destructive techniques, with emphasis to the NIR spectroscopy and chemometrical methods of calibration. Food chemistry was, indeed, one of the most benefited fields by the use of NIR spectroscopy for the determination of a series of properties in complex matrices [7–12]. The very first publication on tomato quality parameters is due to Hong and Tsou [13]. After that, Goula and Adamopoulos [14] have determined moisture, sugars, total acidity, salt and protein in tomato products and Jha and Matsuoka [15] have calibrated the acid/brix ratio of various tomato juices. More recently, Pedro and Ferreira [16] have developed a PLS calibration model for determining solids (total and soluble), lycopene and β -carotene, important micronutrients of tomato products [17–19], using NIR spectroscopy. Besides, fast and reliable techniques are of extreme value when shelf-life studies need to be conducted for launching new tomato products into the market [20,21].

The most popular calibration method in chemometrics is the partial least squares (PLS) regression. The main advantage of PLS is that it can deal with strongly correlated data – which is usually the case in spectroscopy – by performing the calibration over latent variables (or factors) that are defined as linear combinations of the original variables. These factors are extracted in such a way that they have the maximum covariance with the property of interest [22,23].

When several dependent variables are available for calibration, two approaches can be used in PLS regression: the properties can be calibrated for one at a time – using the algorithm known as PLS1 – or, alternatively, they can all be calibrated at once (PLS2). In PLS1, one set of factors is needed for each calibration model individually, whilst in PLS2 a single set of factors is extracted. Whilst the factors in PLS2 are obtained in a single execution of the calculation routine, PLS1 usually gives more precise models [22,23]. Nevertheless, PLS2 gives models with similar predictive abilities to PLS1 when the dependent variables are strongly correlated [22,24]. Furthermore, the PLS2 regression is especially advantageous when dependent variables that can be determined quite precisely are put together with properties that can only be quantified less precisely by their reference methods. In this case, PLS2 tends to give calibration models that present, in general, best predictive abilities than those given by the PLS1 procedure [22].

The main objective of this work is to develop a robust analytical procedure for determining, simultaneously and non-destructively, important parameters of industrialized tomato products using NIR spectroscopy and PLS2 regression. The predictive ability of the final PLS2 model is also compared with those obtained by individual PLS1 models for each property.

2. Experimental

Forty-two samples of tomato concentrate products with total solids content ranging from 6.9 to 35.9% (6.8–31.1°Brix, respectively) were collected in Latin America (Brazil and Argentina), the US and Europe (the Netherlands, Spain, Italy and Greece). These samples presented a reasonable range of variation for building suitable calibration models for the properties of interest.

As tomato products are quite sensitive to moulding, the reference analysis and spectra acquisition were performed immediately after opening the samples.

2.1. Reference analyses

Total solids (%) were determined in triplicate by oven-drying (70 °C) using a Fanem EV8 oven (Fanem Co., São Paulo, Brazil) under vacuum (~150 mmHg absolute pressure) produced by an Edwards E2M8 vacuum pump. Approximately, 3 g of sample were weighted in aluminium capsules with 0.6 g of diatomaceous earth (Merck Cellite® 281) and kept in the oven until constant weight (~4 h).

Soluble solids were determined in duplicate at room temperature by using an Abbe bench-top refractometer (American Optical), with 0.1°Brix accuracy.

Total acidity (%citric acid) was quantified using a Metrohm 702 automatic titrator in MET (monotonic equivalence point titration) mode and a combined pH electrode (Metrohm 6.0232.100). Titrating solution was NaOH 0.1 mol L⁻¹ (Merck 1.09141.1000).

Sugars (sucrose, fructose and glucose) were determined in duplicate by using a Shimadzu HPLC (Shimadzu Co., Kyoto, Japan) equipped with a CTO-10A column oven, a Sil-10A automatic injector, SPD-10AV pumps and a RID-6A refractive index detector. Separation was achieved with a Shodex NH₂P-50 4E (5 μ m, 25 cm \times 0.46 cm) column with a Shodex NH₂P-50G column guard (4.6 mm \times 10 mm). Mobile phase was acetonitrile:water (75:25), isocratic at 1 mL min⁻¹. Sigma–Aldrich standards (S8501 for sucrose, D9434 for glucose and F9048 for fructose) were used for building calibration curves. Tomato products were diluted at 5% and clarified with 10 mL of each of the Karenz reagents (K₄Fe(CN)₆ 0.25 mol L⁻¹ and zinc acetate 1 mol L⁻¹). Injection volume was 5 μ L.

2.2. NIR spectra acquisition

A suitable amount of sample was filled into the bottom of a glass Petri dish (Schott 23 755 48 05), and readings were performed in a Büchi NIRLab N-200 spectrometer (Büchi Labortechnik AG, Postfach) equipped with a MSC-100 diffuse reflectance cell with sample rotation system. Besides promoting a better sampling for spectra acquisition, this rotating system also prevents local heating of the sample by the infrared radiation. Three spectra of each sample were collected, each one using 100 scans in the 4000–10,000 cm⁻¹ range (4 cm⁻¹ of resolution), at room temperature (20–25 °C).

2.3. Multivariate calibration

In this work, vectors are represented by bold lower case, matrices by bold upper case, scalars by italic lower case letters and sequences by subscripts.

Spectra were pre-processed in order to remove or reduce noise, offset and/or bias, where offset was defined as a constant drift and bias as constant slope along the baseline [22,24]. For reducing noise and offset, a mean-smoother with windows

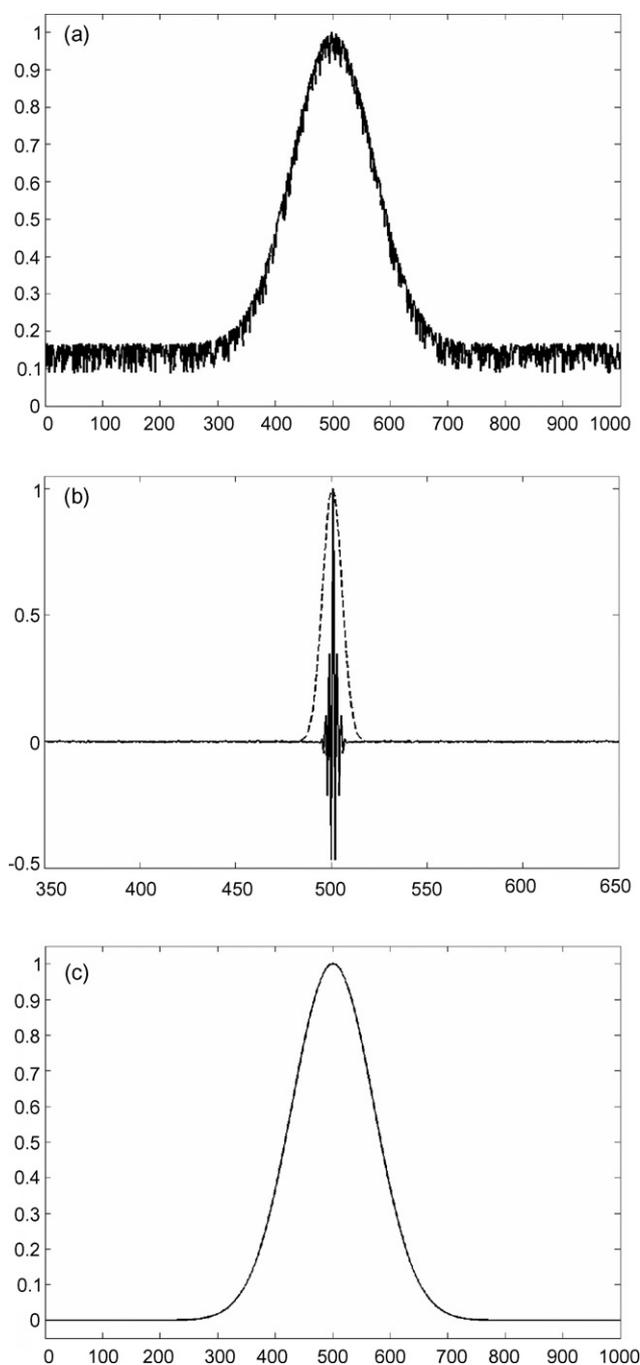


Fig. 1. Fourier smoothing of a generic signal with arbitrary units: (a) noisy peak; (b) a Gaussian convolution function (dashed line) is applied to the interferogram (solid line) in order to remove noisy high-frequency features; (c) smoothed peak.

of different sizes followed by multiplicative signal correction (MSC), was applied [22,24]. Alternatively, a Fourier filter with Gaussian convolution functions of several widths (w in Eq. (1)) was applied (Fig. 1), followed by MSC and first- or second-derivatives according to the algorithm described by Savitzky and Golay [22,24,25].

$$f(x) = e^{[-(x-x)^2/w^2]} \quad (1)$$

The two partial least squares approaches (PLS2 and PLS1) were used for building calibration models and the Y-block was

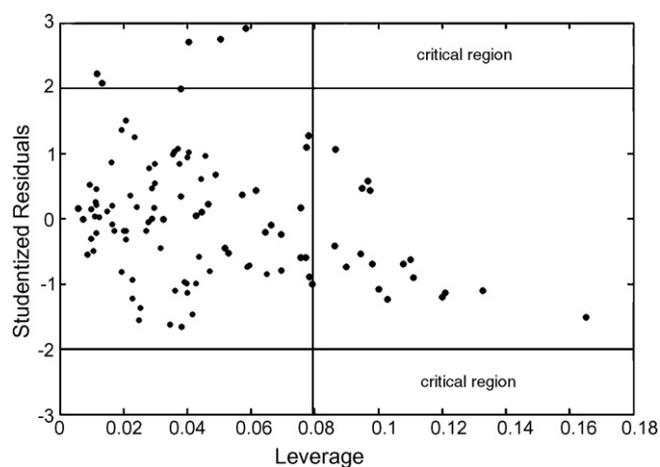


Fig. 2. Leverage vs. studentized residuals chart. The horizontal lines represent studentized residuals with 95% significance while the vertical line represents the critical leverage value.

autoscaled in the PLS2 calibrations. Cross-validation following the leave-one-out procedure was performed in order to define the optimum number of factors to be kept in the models (through RMSECV). For detecting outliers, leverage (Eq. (2)) versus studentized residuals (Eq. (4)) charts (Fig. 2) were built. Samples presenting high-studentized residuals and leverage values above a critical value (Eq. (5)) were considered outliers and removed from the calibration set [22–24,26].

$$h_i = \frac{1}{N_c} + (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{X}^T \mathbf{X})^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})^T \quad (2)$$

$$L_{resc_i} = \sqrt{\frac{(y_i - \hat{y}_i)^2}{(N_c - 1)(1 - h_i)^2}} \quad (3)$$

$$SR_i = \frac{y_i - \hat{y}_i}{L_{resc_i} \sqrt{(1 - h_i)}} \quad (4)$$

$$h_{crit} = \frac{3k}{N_c} \quad (5)$$

where h_i corresponds to the leverage value of the i -th sample; \mathbf{T} to the scores matrix for the whole calibration set; \mathbf{t}_i to the vector of scores of object i ; N_c to the number of samples in the calibration set; L_{resc_i} corresponds to the residual for the i -th sample, standardized by its leverage value; SR_i stands for the studentized residual; y_i and \hat{y}_i are, respectively, the measured and the estimated values of property y for the i -th sample; h_{crit} corresponds to the critical leverage; and k is the number of factors or latent variables.

Seven samples were separated for external validation and the other 35 constituted the calibration set. The PLS2-model with the best predictive ability was selected according to the pooled-RMSEP (RMSEP_p, refer to Eq. (6)) and the pooled r_{val} (Eq. (7)) [22–24]:

$$RMSEP_p = \sqrt{\frac{\sum_{j=1}^J \sum_{i=1}^{N_v} (y_{ij} - \hat{y}_{ij})^2}{J \times N_v}} \quad (6)$$

$$r_{\text{val,p}} = \frac{\sum_{j=1}^J \sum_{i=1}^{N_v} (y_{ij} - \hat{y}_{ij})^2}{\sqrt{s^2(\mathbf{Y}) \times s^2(\hat{\mathbf{Y}})}} \quad (7)$$

in which RMSEP_p is the pooled root mean squared error of prediction, $r_{\text{val,p}}$ the pooled correlation coefficient between the estimated and the predicted values, $s^2(\mathbf{Y})$ and $s^2(\hat{\mathbf{Y}})$, respectively, are the pooled variances of the measured and predicted values for the properties under study, N_v is the number of samples in the validation set and J is the number of properties being calibrated by the PLS2 regression.

The performance of the models obtained with the PLS1 regression was assessed using the RMSEP and r_{val} , where the summations in j in Eqs. (6) and (7) were dropped [22–24].

3. Results and discussion

Fig. 3a shows the original spectra collected for samples presenting low (solid lines) and high (dashed lines) amounts of tomato solids. Considerable noise can be seen in the regions between 4000–5500 and 6300–7100 cm^{-1} , especially in the spectra of samples having low amounts of tomato solids. For removing noise, average-smoothing and Fourier filters were applied.

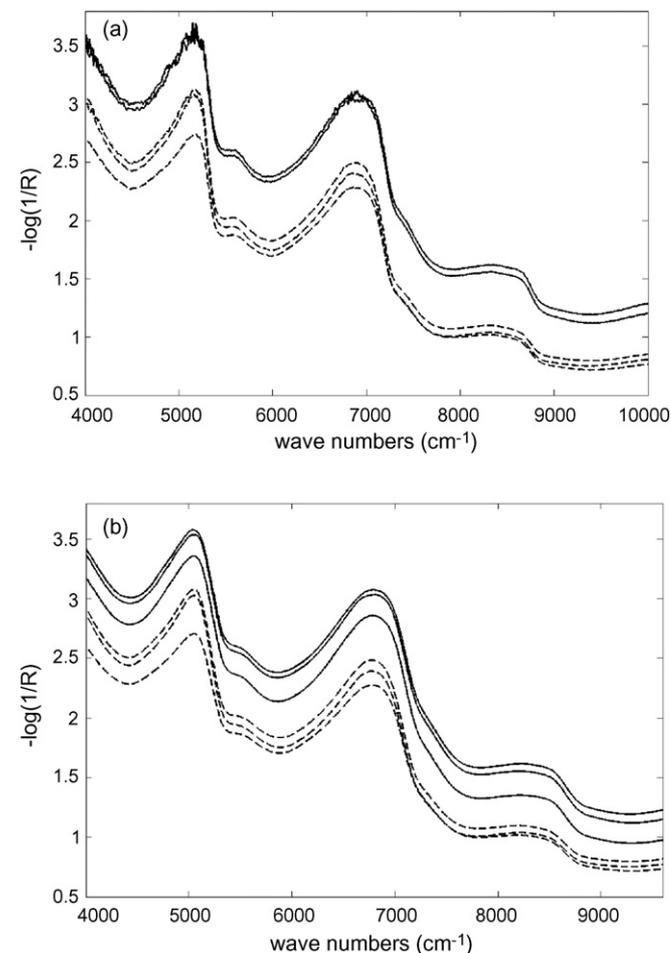


Fig. 3. (a) Original spectra and (b) smoothed spectra. The solid and dashed lines represent, respectively, samples with low and high amounts of tomato solids.

The best window for average-smoothing was determined to be of 50 points (Fig. 3b) in the sense that it gave models with better predictive abilities than those built using other windows. It could be observed that below this limit the spectra were still presenting noise, whereas above 50 points their original shapes were considerably changed. Regarding the Fourier-filter smoothing, a width of 17 points produced the best PLS2 calibrations, the arguments being similar to those stated for the average-smoothing procedure.

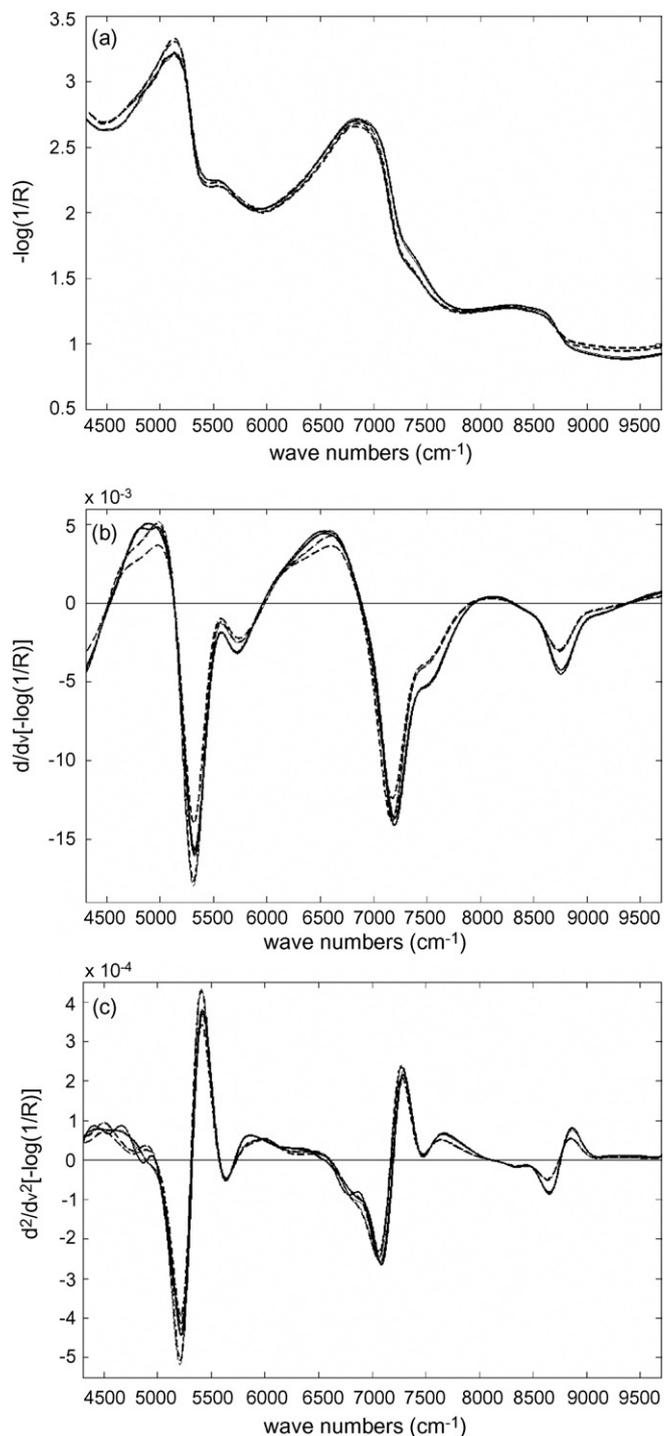


Fig. 4. Spectra pre-processing: (a) MSC; (b) first- and (c) second-derivative after 17-points Fourier-filter smoothing.

Table 1
Predictive abilities and number of factors of the PLS2 models

| Pre-Processing | | # of Factors | RMSECV | RMSEP _p | <i>r</i> _{val,p} |
|-------------------|----------------------------|--------------|--------|--------------------|---------------------------|
| Smoothing | Offset and/or bias removal | | | | |
| | None | 8 | 1.11 | 1.09 | 0.970 |
| | MSC | 5 | 0.89 | 0.94 | 0.979 |
| | 1st-derivative | 9 | 1.32 | 1.35 | 0.957 |
| Average smoothing | None | 8 | 0.88 | 0.89 | 0.981 |
| | MSC | 4 | 0.83 | 0.83 | 0.982 |
| | 1st-derivative | 9 | 1.06 | 1.19 | 0.971 |
| | 2nd-derivative | 9 | 2.12 | 2.15 | 0.856 |
| Fourier filter | None | 8 | 0.84 | 0.83 | 0.982 |
| | MSC | 4 | 0.82 | 0.83 | 0.982 |
| | 1st-derivative | 7 | 0.99 | 0.93 | 0.978 |
| | 2nd-derivative | 7 | 1.08 | 1.00 | 0.971 |

Consistent baseline offsets and bias were present in the spectra as well. These are quite common features in NIR spectra acquired by diffuse reflectance techniques [27,28]. Nevertheless, it was noticed that the offset presented a significant correlation with the amount of solids in the samples (average $r = 0.81$). This might be due to the summation of (a) differences in the path length of the infrared radiation, because, the lower the concentration of particulate material, the deeper the light penetrated into the sample, and (b) due to differences in the scatter profile of the samples and the reference during the reflectance measurements. However, whatever the causes are, these offset and bias usually produce calibration models which require a larger number of factors or with lower predictive abilities (higher RMSEP), and thus, it is usually good practice in chemometrics to remove or reduce these features.

Fig. 4a shows that MSC has effectively reduced the offset originally present in the spectra, but bias was still present. For removing bias, derivatives were applied after Fourier-filter smoothing: while the first-derivative completely removes offset—it only transforms the bias into a constant term (notice, in Fig. 4b, that the baseline in the 9000–9600 cm^{-1} region does not tend to zero), the second-derivative removes both, offset and bias. However, as the derivatives are numerically calculated as differences of the intensities in (usually constant) intervals of wave numbers, they have the detrimental effect of reducing the signal-to-noise ratio, especially when applied to noisy spectra [22,24]. This effect can be seen in Fig. 4c, which shows the second-derivative of the Fourier-filter smoothed spectra.

Table 1 shows the PLS2 models which presented the best predictive abilities (RMSEP_p and $r_{\text{val,p}}$) within each pre-processing technique. No outliers were found in any calibration. The models which presented the lowest RMSEP_p and highest $r_{\text{val,p}}$ values were obtained by pre-processing the spectra using MSC after smoothing, both techniques – average-smoothing and Fourier filter – giving similar results.

Nevertheless, despite both smoothing procedures increase considerably the signal-to-noise ratio without reducing the amount of information in the spectra, Fig. 5 shows that the Fourier filter was more effective on smoothing the spectra than

the average-smoothing technique, and this difference was particularly relevant for the calibrations built with spectra using derivatives as pre-processing. This was due to the fact that, while the former largely reduces the high frequency signals from the interferogram itself, the latter is only based on averages of the (approximately) randomly distributed spectral points in the noisy regions.

The models obtained with derivatives were more complex – in the sense that a larger number of factors were required for building up the calibration models – and presented lower predictive abilities than those obtained with spectra pre-processed by MSC. Despite the Fourier filter being more effective on reducing noise from the original spectra, more factors were needed for accommodating the noise re-inserted into the spectra by the derivatives. This result is in accordance with those previously reported [16], although a different smoothing procedure was applied.

Calibration models were built for each property individually using PLS1 regression as well. Table 2 shows a comparison between the best model obtained using PLS2 (50-point average

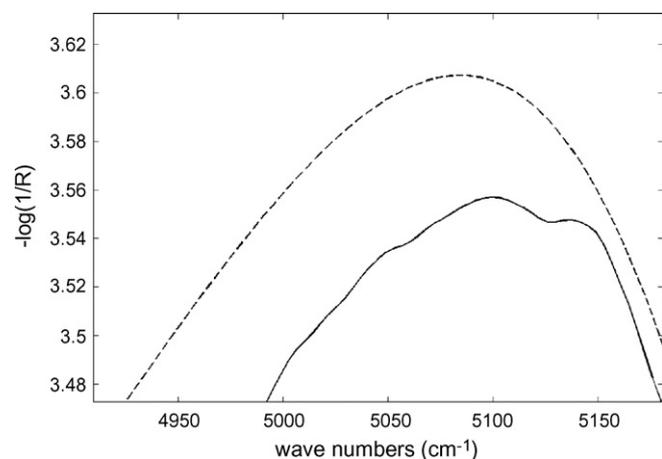


Fig. 5. Differences of the smoothing procedures. The dashed and solid lines represent, respectively, the same Fourier-filtered and average-smoothed spectrum. The dashed line was shifted for clarity.

Table 2
Comparison between models with the best predictive abilities built using PLS2 and PLS1

| Property | PLS2 (four factors) | | PLS1 | | |
|------------------------------|---------------------|------------------|---------|-------|------------------|
| | RMSEP | r_{val} | Factors | RMSEP | r_{val} |
| Total solids (%) | 0.63 | 0.999 | 8 | 0.81 | 0.998 |
| Soluble solids (%) | 0.68 | 0.999 | 9 | 0.35 | 0.999 |
| Total acidity (%citric acid) | 0.22 | 0.970 | 10 | 0.21 | 0.974 |
| Total sugar (%) | 1.96 | 0.962 | 6 | 2.19 | 0.943 |
| Glucose (%) | 0.54 | 0.991 | 6 | 0.59 | 0.990 |
| Fructose (%) | 0.88 | 0.973 | 4 | 0.93 | 0.967 |

Table 3
Correlation coefficients of the properties under study

| | Total solids | Soluble solids | Total acidity | Total sugars | Glucose | Fructose |
|----------------|--------------|----------------|---------------|--------------|---------|----------|
| Total solids | 1.000 | 0.995 | 0.796 | 0.746 | 0.916 | 0.897 |
| Soluble solids | | 1.000 | 0.791 | 0.750 | 0.902 | 0.897 |
| Acidity | | | 1.000 | 0.550 | 0.646 | 0.624 |
| Total sugars | | | | 1.000 | 0.750 | 0.859 |
| Glucose | | | | | 1.000 | 0.935 |
| Fructose | | | | | | 1.000 |

smoothing followed by MSC) and those which presented the best predictive abilities using PLS1 for each property. It can be seen that PLS2 presented better predictive abilities for four out of the six properties under study, namely total solids, total sugar, sucrose and fructose. PLS1 yielded models with better predictive abilities for soluble solids and total acidity, but with models which demanded a larger number of factors.

PLS2 regression resulted in models with better predictive abilities than most of the PLS1 models because of a combination of features of this specific data set. Firstly, the properties under study were strongly correlated, as shown in Table 3 and, secondly, all variables presented high correlation with total solids,

a property that could be determined quite accurately by its reference method.

Table 4 shows the measured and predicted values, as well as the standard errors of prediction for the seven samples used in the external validation. The errors are within those obtained by the reference methods for total and soluble solids and are comparable with results previously reported for NIR analysis [16]. However, they are slightly higher for sugars and considerably higher for total acidity. As total acidity was determined by potentiometric titration, its standard error is usually smaller than 5%. The standard errors for sugar are usually within 10–15%, due to the sample preparation step. The larger error of prediction

Table 4
Measured and predicted values, and standard errors of prediction (SEP) for the individual samples of the validation set

| Sample | Total solids (%) (calibration range: 6.94–35.96%) | | | Soluble solids (%) (calibration range: 6.8–31.1%) | | | Total acidity (%) (calibration range: 0.28–2.46%) | | |
|--------|---|-----------|---------|---|-----------|---------|---|-----------|---------|
| | Measured | Predicted | SEP (%) | Measured | Predicted | SEP (%) | Measured | Predicted | SEP (%) |
| 1 | 8.23 | 7.92 | 3.72 | 7.8 | 7.8 | 0.39 | 0.36 | 0.32 | 11.11 |
| 2 | 19.32 | 19.72 | 2.06 | 18.3 | 17.9 | 1.83 | 0.78 | 0.80 | 2.56 |
| 3 | 16.92 | 16.72 | 1.19 | 16.8 | 16.7 | 0.21 | 0.51 | 0.57 | 11.64 |
| 4 | 19.97 | 19.82 | 0.77 | 19.1 | 19.5 | 2.34 | 1.20 | 1.30 | 9.16 |
| 5 | 25.74 | 27.13 | 5.39 | 25.4 | 25.5 | 0.34 | 1.26 | 1.10 | 12.39 |
| 6 | 9.69 | 9.86 | 1.69 | 9.3 | 9.4 | 0.84 | 0.66 | 0.59 | 10.66 |
| 7 | 14.76 | 14.19 | 3.87 | 14.0 | 13.7 | 2.09 | 0.88 | 0.79 | 9.71 |
| Sample | Total sugars (%) (calibration range: 0.87–17.73%) | | | Fructose (%) (calibration range: 0.92–9.55%) | | | Glucose (%) (calibration range: 1.30–8.18%) | | |
| | Measured | Predicted | SEP (%) | Measured | Predicted | SEP (%) | Measured | Predicted | SEP (%) |
| 1 | 4.45 | 5.41 | 21.60 | 1.71 | 1.52 | 10.98 | 1.34 | 1.46 | 9.06 |
| 2 | 4.89 | 6.00 | 22.70 | 3.96 | 3.06 | 22.74 | 3.39 | 2.77 | 18.31 |
| 3 | 4.12 | 3.08 | 25.24 | 3.15 | 3.22 | 2.26 | 2.67 | 2.22 | 16.87 |
| 4 | 4.75 | 4.09 | 13.80 | 5.22 | 4.43 | 15.20 | 3.07 | 3.48 | 13.44 |
| 5 | 13.99 | 16.40 | 17.23 | 6.73 | 6.09 | 9.53 | 7.26 | 8.51 | 17.21 |
| 6 | 4.18 | 3.20 | 23.40 | 1.78 | 1.95 | 9.78 | 1.61 | 1.49 | 7.44 |
| 7 | 3.24 | 3.46 | 6.86 | 3.33 | 3.70 | 11.13 | 2.63 | 2.32 | 11.97 |

for total acidity is probably due to the fact that this property is indirectly being calibrated using the NIR spectra.

Suitable calibration models could not be obtained for sucrose because this sugar is not given by the tomato itself—it is used as part of the formulation for adjusting the taste of the final industrialised products. Because, the spectra did not present many features, it was observed during this study that only properties which were strongly correlated with the tomato solids in the product could be reasonably calibrated for, but this observation has to be confirmed in a study with a larger number of samples.

4. Conclusion

A suitable PLS2 calibration model was obtained for determining solids, acidity and sugars (total sugars, glucose and fructose) by NIR spectroscopy.

Despite the standard errors of prediction for acidity and sugars being higher than those obtained by the reference methods, the calibration models for these properties could be improved by (a) adding more samples into the calibration set and (b) building up region- or country-specific calibration models.

As tomato products are extremely season-dependent, besides increasing the samples in the calibration set, future work should focus on validating and/or adjusting the calibration models for the between-season variability.

Acknowledgement

The authors would like to thank Mr. Colin Haine, head of the Tomato Global Technology Centre, Unilever, for kindly providing the samples for this study.

References

- [1] W.A. Gould, *Tomato Production, Processing & Technology*, 3rd ed., CTI Pub, Baltimore, 1992, pp. 15–120, 236–349.
- [2] World Processing Tomato Council (WPTC), release # 12, <http://www.wptc.to/releases/releases12.pdf> (accessed 5.10.06).
- [3] USDA Foreign Agricultural Service (FAS), *World Tomato and Tomato Products Situation and Outlook*, http://www.fas.usda.gov/ftp/Hort_Circular/2005/08-05/08-01-05%20Tomato%20article.pdf (accessed 5.10.06).
- [4] D.K. Salunkhe, S.S. Kadam (Eds.), *Handbook of Vegetable Science and Technology*, Marcel Dekker, New York, 1998, pp. 171–2002.
- [5] S.S. Nielsen, *Food Analysis*, 3rd ed., Gaithersburg Pub, Aspen, 2003, pp. 112–132.
- [6] J.L. Cháves-Servín, A.I. Castellote, M.C. López-Sabater, *J. Chromatogr. A* 1043 (2004) 211–215.
- [7] J.Y. Chen, C. Iyo, S. Kawano, *J. Near Infrared Spectrosc.* 7 (1999) 265–273.
- [8] J. Lammertyn, *Trans. ASAE* 41 (1998) 1089–1094.
- [9] S.N. Jha, S. Matsuoka, *Food Sci. Technol. Res.* 6 (2000) 248–251.
- [10] S. Kawano, T. Fujiwara, M. Iwamoto, *J. Jpn. Soc. Hortic. Sci.* 62 (1993) 465–470.
- [11] S.D. Osborne, R. Kunnemyer, R.B. Jordan, *J. Near Infrared Spectrosc.* 7 (1999) 9–15.
- [12] S. Kawano, H. Watanabe, M. Iwamoto, *J. Jpn. Soc. Hortic. Sci.* 61 (1992) 445–451.
- [13] T.L. Hong, S.C.S. Tsou, *J. Near Infrared Spectrosc.* 6 (1998) A321–A324.
- [14] A.M. Goula, K.G. Adamopoulos, *J. Near Infrared Spectrosc.* 11 (2003) 123–136.
- [15] S.N. Jha, T. Matsuoka, *Intern. J. Food Sci. Technol.* 39 (2004) 425–430.
- [16] A.M.K. Pedro, M.M.C. Ferreira, *Anal. Chem.* 77 (2005) 2505–2511.
- [17] P. Bramley, *Phytochemistry* 54 (2000) 233–236.
- [18] H. Tapiero, D.M. Townsend, K.D. Tew, *Biomed. Pharmacother.* 58 (2004) 100–110.
- [19] J. Willcox, G.L. Catignani, S. Lazarus, *Crit. Rev. Food Sci. Nutr.* 43 (2003) 1–18.
- [20] T.P. Labuza, M.K. Schmidl, *Food Technol.* 39 (1985) 57–64.
- [21] A.M.K. Pedro, M.M.C. Ferreira, *Multivariate accelerated shelf-life testing*, *J. Chemomet.* 20 (2006) 76–83.
- [22] H. Martens, T. Naes, *Multivariate Calibration*, John Wiley & Sons, Chichester, 1993, pp. 116–165, 237–290, 314–354.
- [23] R.G. Brereton, *Chemometrics: Data Analysis for the Laboratory and Analytical Plant*, Wiley & Sons, Chichester, 2003, pp. 271–315.
- [24] D.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. de Jong, P.J. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometrics, Part B*, Elsevier, Amsterdam, 1998, pp. 349–370.
- [25] A. Savitzky, M. Golay, *Anal. Chem.* 36 (1964) 1627.
- [26] M.M.C. Ferreira, A.M. Antunes, M.A. Melgo, P.L.O. Volpe, *Quim. Nova* 22 (1999) 724–731.
- [27] P. Williams, K. Norris (Eds.), *NIR Technology in the Agricultural and Food Industries*, Am. Ass. Cereal Chemists, St. Paul, 1990, pp. 1–31.
- [28] J.M. Olinger, P.R. Griffiths, *Appl. Spectrosc.* 47 (1993) 695–701.